Decoding Acute Pain with Combined EEG and Physiological Data

Jenessa Lancaster, Hiroaki Mano, Daniel Callan, Mitsuo Kawato, Ben Seymour.

Abstract— **Across neuroscience research, clinical diagnostics, and engineering applications in pain evaluation and treatment, there is a need for an objective measure of pain experience and detection when it occurs. This detector should be reliable in real-world settings using easily accessible, non-invasive data sources. We present a simple yet robust paradigm for decoding pain using neural and physiological data including electroencephalography (EEG), pulse, and skin conductance (GSR) measurements. The present study uses multivariate classification to distinguish painful events from non-painful multimodal sensory stimuli. To classify the pain response and detect relevant data attributes, we employed a sparse logistic regression (SLR) machine learning protocol with automatic feature selection. EEG input consisted of time-frequency changes under trial conditions, and physiological data included fluctuations and spikes in pulse and skin conductance. Classification averaged 70% accuracy and selected between 5 and 15 features. In our experiment, pain was induced by cold stimulation which became noxious with prolonged exposure. Due to the long, ramp-and-hold nature of the stimulus, along with individual variability in sensitivity to pain, we did not observe specific rapid evoked responses or time-locked events common across participants. However, this format more closely resembles the experience of pain conditions requiring intervention which could be facilitated by a decoding system. The results illustrate the feasibility of developing a wireless pain detection system and give insight to important temporal, spectral, and spatial EEG events and physiological indicators of pain states. Success of the classifier protocol using these parameters could lead to the creation of a closed-loop system for decoding and intervention which can be applied in engineering and medical contexts.**

I. INTRODUCTION

The desire for an objective measure of pain experience and ability to decode pain perception spans the fields of neuroscience, clinical diagnostics and management, and engineering applications for communication and treatment [1]. Due to the need for easily accessible data sources, EEG has been widely utilized to capture and investigate representative brain activity in response to different types of pain sensation. The significance of EEG responses and their correlation to physical and self-reported intensity of stimuli has been well established across pain types, including contact heat pain, noxious cold stimulation, and cutaneous laser stimulation [2], [3], [4]. EEG-based decoding of pain perception, and prediction of pain experience and sensitivity has also proven successful in heat pain and laser stimulation. Good predictive accuracy has been demonstrated in intra-individual and cross-subject cases, using multivariate analysis of spatial, temporal, and frequency-domain features of single-trial EEG responses [5], [6]. Autonomic-mediated physiological responses such as skin conductance, heart rate, and pupilometry have also shown the ability to correctly predict heat-evoked pain with performance comparable to neuroimaging methods, especially when combining more than one data source [7], [8].

Brain-related and physiologic responses may represent non-overlapping information regarding the pain response; combining neuroimaging and physiologic data therefore seems promising, and has not previously been fully explored [7]. Previous study designs also have not focused on the distinction between pain processing and reaction to spontaneous events across other sensory modalities [5]. In addition, our cold pain paradigm necessitates a departure from conventional analysis of precise and predictable evoked potentials, instead yielding more complex induced activity which warrants advanced examination. The present study therefore uses 16-channel EEG, skin conductance, and photoplethysmogram to distinguish noxious cold pain from multimodal sensory stimuli using multivariate classification.

II. METHODS

A. Subjects

Fifteen healthy subjects took part in the study. EEG signal quality was satisfactory in all but one subject, who was excluded from further analysis, leaving 14 subjects (10 male, 4 female) aged 21 to 35. All were informed of study procedures and gave informed consent. Experimental protocol was approved by the ATR ethics committee.

B. Setup and Data Recording

During the experiment, skin conductance, pulse, and 16-channel EEG data were recorded. A dry, active EEG cap was used (ActiCAP Xpress, BrainProducts, Germany) with electrodes at FP1, FP2, FC6, FC2, FC1, FC5, Fz, C4, Cz, C3, P4, Pz, P3, O2, Oz, O1; and with reference and ground electrodes at the earlobes. This cap included interchangeable metal electrodes of lengths ranging from 8 to 14 mm. The cap was pre-prepared to a default configuration for quick setup, and the fit point adjusted for each subject. Any electrodes which caused discomfort were shortened to relieve pressure, and electrodes were lengthened in the case of poor scalp contact such as due to head shape or hair texture. Setup was very quick compared to traditional gelled EEG, yet still customized and high-quality.

Research was supported by MEXT SRPBS "Brain Machine Interface Development" program (Japan), the National Institute for Information and Communications Technology (Japan), and the Wellcome Trust (UK).

J Lancaster, M Kawato and B Seymour are at ATR Laboratories, Japan. H Mano, D Callan, M Kawato and B Seymour are at the Center for Information and Neural Networks (CiNet) National Institute of Information and Communications Technology (NICT). J Lancaster is now also at Imperial College London. B Seymour is at the Computational and Biological Learning Lab, Department of Engineering, University of Cambridge. E-mail: j.lancaster16@ic.ac.uk; bjs49@cam.ac.uk.

Pulse and skin conductance sensors were applied on the hands. A photoplethysmogram was placed on one index finger (Blood Pulse Sensor, BrainProducts), while galvanic skin response electrodes (GSR-MR, BrainProducts) were gelled and placed on the inner pad above and below the second knuckle on a finger of the opposite hand. Data was recorded at a sampling rate of 500 Hz using a V-Amp amplifier and captured with BrainVision Recorder software.

C. Stimuli and Experiment Procedure

The experiment consisted of 20 trials each of three stimulus types: noxious cold "Pain"; an innocuous temperature control, "Cool"; and a "Visual" condition, in randomized order. Each stimulus was presented for 30 seconds, with inter-trial intervals randomized between 4 and 6 seconds. Trial presentation was broken into 3 sessions with short breaks between them, during which participants were allowed to move, stretch, and talk. The complete experiment lasted between 40 and 50 minutes. Participants sat in a quiet room, passively experiencing stimuli, and were instructed to look at a fixation cross on a black screen in front of them, which remained as a default visual except when the visual stimulus was presented. Temperature stimulation was created by a thermode (PATHWAY ATS 30x30 mm, Medoc, Israel) applied to the left volar forearm. Moderate noxious cold was induced by a 2°C stimulus over a 30 second duration. The intended result was a ramp-and-hold, tonic pain experience. Under the cool condition, the thermode cooled to 20°C. For the remainder of the experiment, the thermode maintained a baseline temperature of 30°C. The visual stimulus was a stationary checkerboard image displayed on a monitor.

D. Data Preprocessing

EEG data was detrended and filtered (Butterworth zero phase filters, cutoffs 1 to 90 Hz, 48 dB/oct slope; 60 and 120 Hz notch filters). Independent component analysis (ICA) was utilized to remove movement and blink contamination and excessive noise, where artifactual components were inspected and rejected by hand. Data was projected back to EEG space, resulting in ICA-corrected EEG. Pulse data was detrended and filtered (passband 0.05 to 50 Hz; 60 and 120 Hz notch), and downsampled to 125 Hz. Meanwhile, skin conductance data was detrended and filtered (passband 0.5 to 50 Hz; 60 and 120 Hz notch filters), and downsampled to 25 Hz.

E. Feature Preparation

After preprocessing, epochs were extracted for each 30-second trial. Following this, the subsequent protocol acts on the single-trial level, for each subject and each channel. Features were prepared for input to the classifier as follows. In general, EEG features represented changes in spectral activity under the trial conditions, while physiologic input captured global trends and tracked extreme events in skin conductance and pulse traces.

First, EEG trials were baseline corrected with respect to 1-second pre-stimulus epochs. Signals were transformed to the time-frequency domain with a Hamming-window short-time Fourier transform, using a time resolution of 150 ms and a frequency resolution of 1 Hz. The root mean square signal amplitude was computed across sliding time windows of 2.4 s with 50% overlap and compared to that of the pre-stimulus segment, showing the relative change in activity at each frequency for $1-100$ Hz over 24 time windows. These values served as input feature vectors (1x2400) to the classifier for each of the 16 EEG channels.

For pulse data, feature vectors included instantaneous heart rate, and heart rate variability in the frequency domain, represented as a ratio of low to high frequency activity ($0.04-0.15$ Hz, to $0.15-0.40$ Hz [9]). Skin conductance features reflected frequency content, fluctuations, and amplitude changes, including pre-stimulus to trial comparisons (similar to "SCR", "SCL", and "SCG" described in [7]).

F. Classification

Classification was done by sparse logistic regression (SLR) with automatic feature selection [10], for both within-subjects and cross-subjects analyses. A nested 10-fold cross validation design allowed for two stages of classifier development. First, a classifier was trained on single-channel data, giving performance measures for each channel. Channels were ranked both in terms of their average accuracy across folds of cross-validation, as well as by their performance distribution. In order to avoid favoring variable channels with very high performance in some folds but low in others, the reliability that a high accuracy result represented stable prediction above chance level was verified by t-test. Channels with good accuracy were ranked by significance, informed by the p-value of this t-test. For within-subjects analyses, these channel-ranking lists were customized to each individual, while in the cross-subjects case, the classifier was trained with one subject left out, so channel performances reflected aggregate data across the other thirteen. In the second round, only channels with statistical significance at the level of $\alpha = 0.05$ were considered. A new classifier was trained using only these channels to test the hold out group. The final classification result was determined by a consensus of these channels, in a simplified weighted linear opinion pool consensus strategy [11]. Channel contributions consisted of the posterior probability of the chosen class, weighted by both the channel's achieved accuracy in the previous level and the strength of the associated p-value.

III. RESULTS

Subject reports of the pain stimulus varied from very painful to mildly discomforting. Debriefing also suggested significant variability in the time taken for pain to develop after the cold stimulus started.

A. Brain Activity

Between the pain and innocuous temperature conditions, there was an increase in frontal midline high-frequency gamma activity (at FC1, Fz, and Cz, $68-87$ Hz, $p < 0.05$), as well as frontal beta activation (FC1, $18-19$ Hz, $p < 0.05$).

Frequency	Channel	P-Value	Possible Source
57	P4	0.041	Line noise
44	Cz	0.028	Gamma or possible noise
45		0.0421	
50		0.0188	
51		0.0088	
52		0.0214	
84	Cz	0.0334	High-Frequency Gamma
85		0.0087	
86		0.0072	
87		0.0326	
68	Fz	0.046	
69		0.0412	
70		0.0434	
71		0.0375	
72		0.0305	
73		0.0395	
40	O1	0.0366	Gamma
41		0.0405	
18	FC1	0.0498	Evoked 1-20 Hz [6]
19		0.0312	
79	FC1	0.033	High-Frequency Gamma
80		0.0178	

TABLE I

SPECTRAL ACTIVITY SIGNIFICANTLY DIFFERING BETWEEN PAIN AND NON-PAIN EVENT CONDITIONS.

Lower gamma-range fluctuations appeared in central and occipital regions (Cz, O1, 40−52 Hz, $p < 0.05$) detailed in Table I.

Given the variability in subject experience, further evaluation of the pain response was conducted post hoc on an individual basis. Some subjects displayed a clear response spatially, temporally, and spectrally. The major relevant component isolated through ICA in a representative individual shows spatial activation in frontal and central regions, most prominently Fz and Cz (Figure 1a). The time-frequency activation of this component shows peak activity between 1−6 seconds after pain stimulus onset, with maximum power fluctuation in the 1−15 Hz frequency range (Figure 1b).

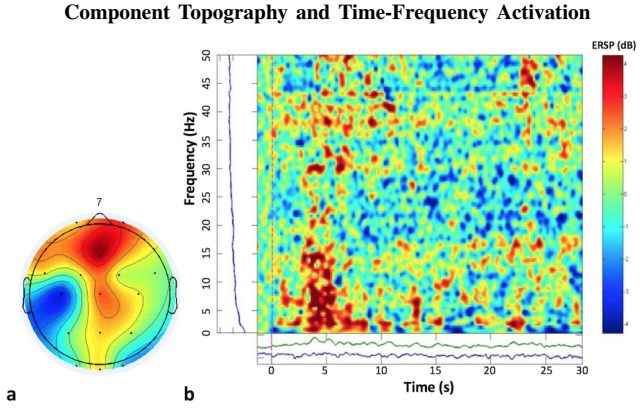

Fig. 1. a) Topography of spatial activation and b) time-frequency representation of a component of the pain response identified by ICA decomposition.

B. Decoding Accuracy

For within-subjects analysis, performance in the final classification scheme reached 79% (cross-participants average 70% ± 5.2%; range 62−79%). In thirteen of fourteen subjects, high performance was verified to represent greater than chance decoding by t-test ($p < 0.05$ in ten-fold cross-validation). This result was also evaluated via permutation statistics. Data was re-run through the classification protocol 1000 times, with true class labels randomly shuffled. The resulting performance distribution across 1000 permutations was compared to that of the real data across 10-fold cross-validation. For 12 of 14 subjects, this showed the true classification accuracy was well above chance ($p < 0.005$).

A cross-subjects analysis conducted in the same manner performed similarly, with a maximum accuracy of 70% (group average 67% ± 1.6%; greater than chance, $p < 0.001$).

In the first stage, single-channel classification consistently showed one or more channels for each subject with good performance (70−86% accuracy). However, abstracting this result to a final single-channel protocol was insufficient, as the identified channel did not necessarily perform well on new hold-out test cases. Combining information from a set of several stable channels in a consensus scheme therefore strengthened performance, ensuring high accuracy reliably generalized across training and test cases. Channel sets used in the final classification scheme differed for each subject in within-subjects analysis, and across left-out subjects in the cross-subjects scheme. For almost all subjects, however, the ideal channel set was a combination of EEG and physiologic data. A few runs favored EEG-only analysis, but in no case did physiological data alone outperform the combined sets.

C. Features

Features selected by the classifier represented points in time and frequency for EEG channels. Each classification run yielded between five and fifteen features. Features varied across channels and across subjects in both the time and frequency domains, as well as differing across folds of cross-validation. Aggregating features associated with the pain class which were most commonly selected across subjects revealed that the relevant frequency bands identified in brain activity evaluation were also favored in feature selection in the frontal and central regions of interest. In Cz, the strongest features included 1−20 Hz and 78−98 Hz activity; in Fz, 10−20 Hz, 40−50 Hz, and over 80 Hz; and in FC1, 1−20 Hz and over 70 Hz were selected for.

In some subjects, selected features and their weight magnitudes directly related to actual signal change between baseline and the trial condition. The input vector for a representative subject is shown as a time-frequency representation of relative signal change between trial and pre-trial epochs, for pain with the pain-absent case subtracted, and with feature weight magnitudes overlaid (Figure 2).

IV. DISCUSSION

This study demonstrates that the pain response can be reliably distinguished from general sensory stimulus responses. Good binary classification was achieved despite inconsistencies in the pain class, as well as the combination of unlike non-pain events in the pain-absent class. Cold, tonic pain created by a moderate intensity over a long duration was decodable with simple analysis techniques. The combination

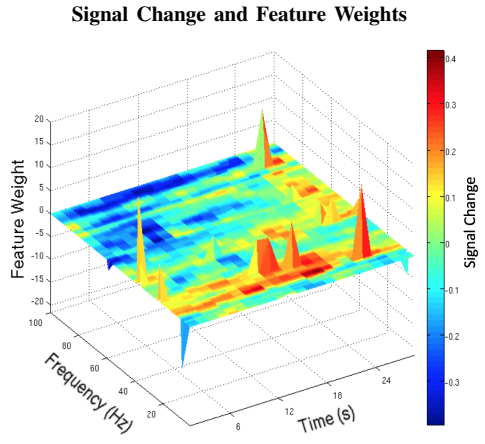

Fig. 2. Selected features and their weight magnitudes, and their relationship to actual signal change between baseline and the trial condition.

of EEG with physiologic data sources allowed the highest-accuracy performance, comparable to or exceeding EEG-only or physiologic-only protocols established in previous literature. This approach, allowing channel set customization while employing a weighted consensus strategy which penalized instability, improved testing accuracy and usability.

Due to individual variation in sensitivity, the pain stimulus was perceived by some as intense and others as only mild. Considering the nature of this stimulus, whereby pain was induced increasingly over prolonged exposure, subjects may each have felt it became painful at varying times within each 30-second trial, and onset times may have differed across trials within subjects. Response patterns in spatial and time-frequency domains reflected this variation, and lost definition in group averages, encouraging the present individual-focused analysis. Future studies may add a calibration step to standardize the intensity and timing of pain induction, or explore continuous decoding of a variety of pain levels.

Self-reported measures of pain experience were not collected, so for classification purposes, the entire 30-second epoch of 2° stimulation was considered a 'pain' event. This likely also contributed to cross-subject differences in feature selection in time and frequency. Several iterations of classifier construction, paired with the SLR method constraints favoring sparseness, collapsed the time series into few representative features, which may have fluctuated across runs. This limited interpretability of selected features, but allowed for decoding of a rich feature set which was quite large relative to the number of training trials. Aggregating all chosen features positively associated with the pain class revealed agreement with conventional analyses. Relevant brain activity bands, especially $1-20$ Hz and high-frequency gamma, were represented in the feature set with a higher occurrence. Though high gamma could originate from artifact such as muscle activity, ICA component spatial locations and signal patterns were inspected to identify and remove those arising from movement. Other recent studies have identified similar activity in these frequency ranges as meaningful components of the pain response [6].

This paradigm presented the challenge of allowing potentially large variability into the target decodable trials, but also demanded particular robustness of the analysis procedure to overcome it. The success of the end classifier protocol demonstrates promise that this method may hold up in under-defined circumstances, which may prove promising for real-world applications in which noise, inconsistency, and minimally controlled event parameters are inevitable. Within-subjects analysis performed well despite a lack of coherent signal presentation or clear prior expectation of signal morphology. This suggests that an individually customized approach could lead to robust classification within subjects, even when little structure is imposed upon the experimental pain experience. This protocol provides a basis for further development of a system which could perform reliably in a realistic adverse pain scenario, and the minimally invasive approaches used could be developed into convenient sensor technology. The potential applications of this achievement range from clinical diagnosis, such as improved communication of symptoms, to neuroengineering, such as automatic dosage regulation in electrical stimulation-based pain relief devices. Further preliminary study would be useful in characterizing the pain complex and its manifestations in EEG, especially in tonic or chronic pain, and more extensive characterization of patient experience could be helpful in guiding feature selection and tuning classification protocols.

References

[1] S. Zhang and B. Seymour, "Technology for chronic pain," *Current Biology*, vol. 24, no. 18, pp. R930–R935, 2014.

[2] A. C. Chen, D. M. Niddam, and L. Arendt-Nielsen, "Contact heat evoked potentials as a valid means to study nociceptive pathways in human subjects," *Neuroscience letters*, vol. 316, no. 2, pp. 79–82, 2001.

[3] S. Shao, K. Shen, K. Yu, E. P. Wilder-Smith, and X. Li, "Frequency-domain eeg source analysis for acute tonic cold pain perception," *Clinical Neurophysiology*, vol. 123, no. 10, pp. 2042–2049, 2012.

[4] L. Garcia-Larrea, M. Frot, and M. Valeriani, "Brain generators of laser-evoked potentials: from dipoles to functional significance," *Clinical neurophysiology*, vol. 33, no. 6, pp. 279–292, 2003.

[5] G. Huang, P. Xiao, Y. Hung, G. D. Iannetti, Z. Zhang, and L. Hu, "A novel approach to predict subjective pain perception from single-trial laser-evoked potentials," *Neuroimage*, vol. 81, pp. 283–293, 2013.

[6] E. Schulz, A. Zherdin, L. Tiemann, C. Plant, and M. Ploner, "Decoding an individual's sensitivity to pain from the multivariate analysis of eeg data," *Cerebral Cortex*, p. bhr186, 2011.

[7] S. Geuter, M. Gamer, S. Onat, and C. Büchel, "Parametric trial-by-trial prediction of pain by easily available physiological measures," *PAIN®*, vol. 155, no. 5, pp. 994–1001, 2014.

[8] R. Treister, M. Kliger, G. Zuckerman, I. G. Aryeh, and E. Eisenberg, "Differentiating between heat pain intensities: the combined effect of multiple autonomic parameters," *PAIN®*, vol. 153, no. 9, pp. 1807–1814, 2012.

[9] G. A. Reyes del Paso, W. Langewitz, L. J. Mulder, A. Roon, and S. Duschek, "The utility of low frequency heart rate variability as an index of sympathetic cardiac tone: a review with emphasis on a reanalysis of previous studies," *Psychophysiology*, vol. 50, no. 5, pp. 477–487, 2013.

[10] O. Yamashita, M. Sato, T. Yoshioka, F. Tong, and Y. Kamitani, "Sparse estimation automatically selects voxels relevant for the decoding of fmri activity patterns," *NeuroImage*, vol. 42, no. 4, pp. 1414–1429, 2008.

[11] G. J. Briem, J. A. Benediktsson, and J. R. Sveinsson, "Multiple classifiers applied to multisource remote sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2291–2299, 2002.